Theoretical Computer Science 64 (1989) 203-219
North-Holland

EXACT SOLUTIONS FOR OPEN, CLOSED AND MIXED QUEUEING NETWORKS WITH REJECTION BLOCKING

I.F. AKYILDIZ*

School of Information and Computer Science, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

H. von BRAND

Departemento Informatica, Universidad Tecnica Federico Santa Maria, Valparaiso, Chile

Abstract. Open, mixed and closed queueing networks with multiple job classes, reversible routing and rejection blocking are investigated in this paper. Jobs may change class membership and general service requirement distributions that depend on the job class are allowed. We prove that the equilibrium state probabilities have product form if at all stations either the scheduling discipline is symmetric or all service requirements at the station have the same exponential distribution. The solution implies insensitivity in this kind of blocking networks, i.e. the distribution of the jobs in equilibrium, irrespective of their remaining service requirements, depends only on their mean service requirement.

1. Introduction

A *queueing network* is an interconnected collection of *stations*, i.e. devices with queues, in which *jobs* move from one station to the next requesting service. Queueing networks have enjoyed increasing popularity as models of manufacturing computer and communication systems over the last two decades. Jackson [16] and Gordon and Newell [14] show that open and closed networks with one job class, exponential service time distributions and First Come First Served (FCFS) scheduling in all stations have *product form solutions*. This is extended to multiple job classes in open, mixed and closed networks by Kelly [17]. Baskett, Chandy et al. [4] give product form solutions for service requirement distributions that have rational Laplace transforms, represented by stages. Multiple job classes can be accommodated, and the service requirement distributions can depend on the class. Several scheduling disciplines are allowed for stations with nonexponential service requirement distributions. More general scheduling disciplines for this same case are obtained by Kelly [18], whose result is restricted to service requirement distributions that are finite mixtures of Erlang distributions. Barbour [3] proves Kelly's conjecture [18] that the results are applicable to general distributions. For differentiable service requirement distributions Chandy and Martin [8] give necessary and sufficient conditions for product form solutions. The allowable scheduling disciplines turn out to be

* Supported in part by the Air Force Office of the Scientific Research (AFOSR) under Grant AFOSR-87-0160.

exactly those for which Kelly [18] and Barbour [3] show product form solution. For short, we call the above types of networks *classical.* Efficient exact algorithms to compute performance measures are known for classical networks: Convolution [6, 22, 27], Mean Value Analysis (MVA) [29], LBANC [10].

In many situations of practical interest, under certain circumstances a station will refuse a job arriving at it. This phenomenon is called *blocking.* Many different blocking policies have been considered in the literature. Onvural and Perros [25] compare different blocking policies and derive equivalencies between them.

The organization of the rest of the paper is as follows: Section 2 gives an overview of our model and a short survey of queueing networks with rejection blocking. Section 3 gives definitions and our notation for queueing networks, and completes the description of the model given in Section 2. In Section 4 we prove the product form solution for the equilibrium state probabilities, and Sections 5 and 6 contain the proofs of some consequences of the equilibrium state distribution given in Section 4.

2. The rejection blocking policy

We consider the so-called rejection blocking policy. Once a job in class α finishes service in station i it determines, according to the routing probabilities $p_{i\alpha,j\beta}$, to which station j and class β it goes next. With a certain probability (that depends on the state of the destination station) the job will be rejected there. The rejected job returns to station i (in class α) to get another round of service, independent from the one it received before. When this new round of service is finished, the job again selects a destination station and class (independent from the ones selected before) and so on.

The rejection blocking policy has the virtue that deadlock is impossible if the network is irreducible, since if there is a free place in some station eventually a job will move into it even if this takes a long sequence of trials. It leads to simple balance equations that are much more tractable than their counterparts for other blocking policies.

The rejection blocking policy was introduced by Caseau and Pujolle [7], who consider tandem networks only. They investigated various blocking policies and general service requirement distributions, with the aim of obtaining bounds on throughput. Pittel [28] showed that rejection blocking models with multiple classes and reversible routing have product form solution if the probability that a job in class α is accepted in station i when there are k_i jobs in station i, of which $k_{i\alpha}$ are of class α, is given by

$$h_{i\alpha}(\boldsymbol{k}_i) = \frac{k_{i\alpha}+1}{k_i+1}\, h_i(k_i). \tag{1}$$

Here h_i is a nonnegative function. Pittel's work is restricted to exponential service requirement distributions and jobs that do not change class membership.

Cohen [11] found a product form solution for a cyclic rejection blocking network with two stations, multiple job classes and class-dependent service requirement distributions (not necessarily exponential) when the scheduling disciplines are processor sharing with load dependent service efforts. The blocking functions allowed are of the form:

$$h_{i\alpha}(\boldsymbol{k}_i) = h_{i\alpha}(k_{i\alpha})h_i(k_i). \tag{2}$$

Here $h_{i\alpha}$ and h_i are nonnegative functions.

Hordijk and van Dijk [15] showed that for special cases of queueing networks with rejection blocking the solutions have product form. They consider models with a single job class in which routing is reversible and models in which blocking is dominant, i.e. there are so many jobs in the system that no station can ever be empty. Balsamo and Iazeolla [2], based on the above work, find classical networks that share part of the state space with rejection blocking networks and show that their equilibrium state probabilities agree (up to a normalization constant) with those of the blocking network on the intersection of the state spaces.

Van Dijk and Tijms [13] give a proof of insensitivity of the distribution of jobs (i.e. dependence only on the mean of the service requirement distribution) in a cyclic network with two stations, multiple job classes and symmetric scheduling disciplines. The blocking functions allowed are of the same form as those allowed by Cohen [11]. Tijms [32, Section 2.5] presents this result in more accessible terms.

Akyildiz and Von Brand [1] prove dualities for rejection blocking networks with one job class and exponential service requirement distributions. There is no restriction on the structure of the network. They show that from a given network one can construct another network of the same general type with the same structure of the state space and such that the throughputs in the given network and its dual are the same. The construction yields an open network if one starts with an open network and a closed network if the given network is closed. Using this result, they are able to prove a product form solution for the case of a closed blocking network in which at most one station can be empty at a time. The solution provides a simple way to compute performance measures, in particular throughputs.

3. Notation and conventions

We consider queueing networks with N stations and C job classes. First we describe an isolated station, and then we turn to describe the interactions between stations in the network.

3.1. An isolated station

A job of class α requests service at station i distributed as $F_{i\alpha}$ with mean $1/\mu_{i\alpha}$. By the results of Barbour [3], it is enough to establish our results for distribution functions that are finite mixtures of Erlang distributions. The restrictions that

Barbour imposes on the network are that there be no multiple transitions and that the arrival processes be independent of the state of the network. Both are satisfied in the model described here.

We will represent the service requirement distributions as mixtures of Erlang distributions of the following form:

$$F_{i\alpha} = \sum_t g_{i\alpha;t} E_{t\nu_{i\alpha}}, \tag{3}$$

where $E_{t\nu_{i\alpha}}$ is the Erlang distribution with t phases, each with rate $\nu_{i\alpha}$. We assume that the sum in (3) is finite, but we refrain from giving the limits to keep the notation simple. Definition (3) means that with probability $g_{i\alpha;t}$ a job of class α arriving at station i will have to traverse t exponential phases, each of which has rate $\nu_{i\alpha}$. This requires

$$\sum_t g_{i\alpha;t} = 1. \tag{4}$$

It also implies

$$\frac{1}{\mu_{i\alpha}} = \sum_t g_{i\alpha;t} \frac{t}{\nu_{i\alpha}}. \tag{5}$$

By renewal theory the probability that at an arbitrary instant a job with service requirement distribution $f_{i\alpha}$ still has to traverse s phases is given by

$$r_{i\alpha}(s) = \frac{\mu_{i\alpha}}{\nu_{i\alpha}} \sum_{t \geq s} g_{i\alpha;t}. \tag{6}$$

Note that

$$r_{i\alpha}(1) = \mu_{i\alpha} / \nu_{i\alpha}. \tag{7}$$

We will denote the state of station i by

$$((\kappa_{i1}, \sigma_{i1}), (\kappa_{i2}, \sigma_{i2}), \ldots, (\kappa_{ik_i}, \sigma_{ik_i})). \tag{8}$$

Here k_i is the number of jobs in station i, κ_{il} is the class of the job in position l of station i and σ_{ij} is the number of remaining phases of service for that job. We will denote the number of jobs of class α in station i by $k_{i\alpha}$.

3.1.1. *Scheduling disciplines*

A scheduling discipline (f, ϕ, ψ) is defined as follows [8, 9, 18, 19]:

- $f(k)$: total service effort when there are k jobs in the station;
- $\phi(l, k)$: fraction of the service effort destined to the job in position l when there are k jobs in the station (zero for l outside of $1 \leq l \leq k$); this requires

$$\sum_{1 \leq l \leq k} \phi(l, k) = 1 \quad \forall k; \tag{9}$$

- $\psi(l, k)$: probability that an arriving job is placed in position l when there are k jobs in the station (zero for l outside of $1 \leq l \leq k+1$); this requires

$$\sum_{1 \leq l \leq k+1} \psi(l, k) = 1 \quad \forall k. \tag{10}$$

Kelly [18] calls a scheduling discipline *symmetric* (Chandy and Martin [8] call them *station balancing*) if

$$\psi(l, k) = \phi(l, k+1). \tag{11}$$

This framework clearly does not describe all possible scheduling disciplines, for example there is no way to give one job class priority over another. Scheduling disciplines that depend on the service requirements, like Shortest Job First (SJF), cannot be described either. Nevertheless, the class of scheduling disciplines that can be described is rich. Some examples are

- FCFS: first come first served is described by $\phi(1, k) = 1$ and $\psi(k+1, k) = 1$;
- LCFS: last come first served preemptive resume is described by $\phi(k, k) = 1$ and $\psi(k+1, k) = 1$;
- PS: processor sharing is described by $\phi(l, k) = 1/k$ and $\psi(k+1, k) = 1$;
- RAND: service in random order [31] is described by $\phi(1, k) = 1$ and $\psi(l, k) = 1/k$ for $l \geq 2$.

Other scheduling disciplines that lead to product form in classical queueing networks, like LBPS (last batch processor sharing, [23]) can also be described [8].

It should be noted that the description of a particular scheduling discipline is not unique. For example, the description for PS given above is *not* symmetric, but if we set $\psi(l, k) = 1/(k+1)$ the discipline becomes symmetric. The only difference between the two is that this alternative does not keep the jobs in their order of arrival, while the description given above does. Of the remaining disciplines, FCFS and RAND are not symmetric, while LCFS is.

We assume that a job selects a service requirement before starting to get service, i.e. when a job enters station i in class α it is assigned a number of phases of service according to the $g_{i\alpha;s}$. If a job in class α is in position l of station i and the number of jobs in station i is k_i, the rate at which that job advances to its next phase of service (or finishes service at the station if it is in its last phase of service there) is $\nu_{i\alpha} f_i(k)\phi_i(l, k_i)$.

3.1.2. *Blocking functions*

We call the probability that a job is accepted at a station the *blocking function* of the station. In the most general case, the blocking function of a station could depend on the state of the entire network. In our model (as in the models of Pittel [28], Hordijk and van Dijk [15, 24, 26], Cohen [11] and van Dijk and Tijms [13]), we allow a dependence only on the state of the destination station. The probability that a job is accepted depends on its class.

Define a partition of the job classes, and denote the set of job classes that contains class α by $[\alpha]$. We write the probability that a job of class α arriving at station i is accepted when there are a total of k_i jobs in it, of which $k_{i\alpha}$ are of class α and $k_{i[\alpha]}$ are of classes in the set that contains class α, as

$$b_{i\alpha}(\boldsymbol{k}_i) = h_{i\alpha}(k_{i\alpha}) h_{i[\alpha]}(k_{i[\alpha]}) h_i(k_i). \tag{12}$$

Here $h_{i\alpha}$, $h_{i[\alpha]}$ and h_i are arbitrary. The only restriction on them is that if $h_i(l)=0$ then $h_i(k)=0$ for all $k \geq l$. Similar restrictions apply to $h_{i\alpha}$ and $h_{i[\alpha]}$. The smallest l as above is then the maximal capacity of the station for jobs, for jobs of class α and for jobs of classes in $[\alpha]$, respectively. These restrictions are needed to ensure irreducibility of the Markov process that represents the queueing network.

More generally, it is possible to take several independent partitions of the job classes and define the blocking function of a job in class α as a product similar to the one in (12) over all partitions (note that we have the partition into single job classes, an arbitrary partition and the partition into a single set in that expression). To divide the job classes into partitions is only a notational convenience, since we can assign $h_{i[\alpha]}(k)=1$ whenever we do not wish jobs in a certain set of classes to be blocked in some partition.

3.2. The network

The state of the network will be described by (ordered) N-tuples of station states. We will use $\boldsymbol{x}$ and $\boldsymbol{y}$ to denote arbitrary states of the network. We define the *occupancy* of the network as an N-tuple of strings of job classes, where the ith string represents the classes of the jobs in station i in order. The *population* of the network gives the numbers of jobs of each class in each station. Occupancies and populations are defined in the obvious ways for single stations. The occupancy of the network will be denoted by $\boldsymbol{n}$, and the population by $\boldsymbol{k}$. For single stations we will use $\boldsymbol{n}_i$ and $\boldsymbol{k}_i$, respectively.

A class α job that tries to leave station i to go to station j in class β but is rejected there returns to station i in class α. It is treated exactly like an arriving job, only that it cannot be rejected. Note that our model differs from the model of Van Dijk and Tijms [13], and of Hordijk and Van Dijk [24, 26] in that they specify that the job returns to the *same position* in station i's queue. In our model it may be placed in any position of the queue, as long as the scheduling discipline allows it.

The structure of the network itself is fixed by the following:

- $p_{i\alpha,j\beta}$: routing probabilities. Probability that a job of class α that leaves station i tries to enter station j in class β. Direct feedback is not allowed, i.e. $p_{i\alpha,i\beta}=0\ \forall i$, α, β.
- $p_{0,j\beta}$: probability that an exogenous job tries to enter station j in class β. We assume that new jobs arrive at a (fixed) rate γ to the network. The process that generates exogenous arrivals is assumed to be Poisson.
- $p_{i\alpha,0}$: probability that a job that finished service in station i in class α leaves the network.

One can define an equivalence relation on pairs (station, job class) by defining $(i,\alpha)\equiv(j,\beta)$ iff a job that starts in station i, class α can wind up in station j, class β after a series of transitions. We call each of the equivalence classes of this relation a *routing chain* or chain for short. Without loss of generality we assume that the sets of job classes in different routing chains are disjoint. So we can identify a routing chain with a set of job classes.

The network is *closed for routing chain* Γ if $p_{0,j\beta}=0$ for all j, $\beta\in\Gamma$. The network is *open for routing chain* Γ if $p_{0,j\beta}$ is non-zero for at least one station j and job class β in routing chain Γ. The network is *closed* if it is closed for all routing chains. The network is *mixed* if it is closed for some routing chains and open for others. The network is *open* if it is open for all routing chains.

For future convenience, we define the $e_{i\alpha}$ by

$$e_{i\alpha}=\gamma p_{0,i\alpha}+\sum_{j\beta} e_{j\beta}p_{j\beta,i\alpha}. \tag{13}$$

There will be one such system of equations for each routing chain. Note that for closed chains the linear system (13) is homogeneous. In that case, we take any particular solution of the system as the $e_{i\alpha}$.

In classical networks the $e_{i\alpha}$ are the throughputs of station i for jobs of class α if the network is open for the routing chain that contains job class α. If the network is closed for the routing chain that contains class α, they can be interpreted as relative throughputs. In the present case these quantities have no physical significance, since the routing of the jobs depends not only on the routing probabilities but also on blocking.

We furthermore assume that routing is *reversible*, i.e.

$$\begin{aligned} e_{i\alpha}p_{i\alpha,j\beta} &= e_{j\beta}p_{j\beta,i\alpha} \quad \forall i,\ j,\ \alpha,\ \beta,\\ \gamma p_{0,j\beta} &= e_{j\beta}p_{j\beta,0} \qquad \forall j,\ \beta. \end{aligned} \tag{14}$$

Reversible routing means that the Markov chain of the pairs (station, class) visited by a job is reversible [19]. In a classical network reversible routing means that the flow of jobs from station i and class α to station j and class β is the same as the flow of jobs from station j and class β to station i and class α. This interpretation is not applicable to blocking networks.

We introduce the following operators:

- $A_{il}(\boldsymbol{x})$: advances the lth job in station i to the next phase of its service requirement (defined whenever $l\le k_i$ and $\sigma_{il}>1$).
- $D_{il}(\boldsymbol{x})$: deletes the lth job in station i (defined only when $l\le k_i$ and $\sigma_{il}=1$). The jobs in positions $l+1, l+2, \dots, k_i$ are shifted forward to positions $l, l+1, \dots, k_i-1$, respectively.
- $I_{il;\alpha s}(\boldsymbol{x})$: inserts a job of class α in the lth position in station i and s phases of service left (defined whenever $l\le k_i+1$). The jobs in positions $l, l+1, \dots, k_i$ (if any) are shifted back one position to $l+1, l+2, \dots, k_i+1$, respectively.
- $T_{ik,kl;\alpha s}(\boldsymbol{x})$: transfers the job in position k of station i to position l of station j and class α with s phases of service left (defined whenever $k\le k_i$, $\sigma_{ik}=1$ and $l\le k_j+\hat{\delta}_{ij}$, where $\hat{\delta}_{ij}$ is defined in terms of Kronecker's delta by $\hat{\delta}_{ij}=1-\delta_{ij}$). The same as $I_{jl;\alpha s}(D_{ik}(\boldsymbol{x}))$.

When discussing the balance equations, we will need the inverses of these operators to describe the state from which the network enters state $\boldsymbol{x}$. Except for the case of $T_{il,jl;\alpha s}$ and D_{il}, the inverse is uniquely defined. When one of these operators are applied, the class in which the affected job was is lost. We will not need an inverse for D_{il}. We write $T_{il,jt;\alpha s}^{-1(\beta)}$ for the inverse of $T_{il,jt;\alpha s}$ if the affected job comes from class β in what follows.

4. The equilibrium state distribution

Let S be the set of feasible states of the network, i.e. states in which the capacity of no station is exceeded, and $q(x, y)$ the transition rate from state x to state y. The global balance equations can then be written:

$$\pi(x) \sum_{y \in S} q(x, y) = \sum_{y \in S} q(y, x) \pi(y). \tag{15}$$

For later convenience, we define

$$q(x) = \sum_{y \in S} q(x, y) \tag{16}$$

That is, $q(x)$ is the total rate out of state x.

To keep the equations readable, we will assume that the job at position l of station i is in class κ and has σ phases of service left. With the above notation and these conventions, we can write down the transition rates from state x to other states as follows:

$$q(x, I_{jm;\beta s}(x)) = p_{0,j\beta} \gamma b_{j\beta}(k_j) \psi_j(m, k_j) g_{j\beta;s}, \tag{17}$$

$$q(x, D_{il}(x)) = p_{i\kappa,0} \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i), \tag{18}$$

$$q(x, A_{il}(x)) = \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i), \tag{19}$$

$$q(x, T_{il,jm;\beta s}(x)) = p_{i\kappa,j\beta} \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i) b_{j\beta}(k_j) \psi_j(m, k_j) g_{j\beta;s}, \tag{20}$$

$$q(x, T_{il,im;\kappa s}(x)) = \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i) \psi_i(m, k_i - 1) g_{i\kappa;s} \cdot \sum_{j\beta} p_{i\kappa,j\beta}(1 - b_{j\beta}(k_j)). \tag{21}$$

Equation (17) corresponds to exogenous jobs entering the network while (18) corresponds to jobs leaving the network. Equation (19) is for a job that finishes a phase of its service and advances to the next one. Equations (20) and (21) are for jobs that try to go from station i to station j, successfully in (20) and unsuccessfully in (21). No other transitions are possible. If the network is closed, the transitions described by (17) and (18) are also ruled out.

Now we can state our principal result, the following theorem.

Theorem 1. *Consider an open, closed or mixed queueing network with rejection blocking in which routing is reversible and there is no direct feedback. Assume that all stations satisfy one of the following*:

(i) *They have symmetric scheduling disciplines with general service requirement distributions that may depend on the job class. We call these stations type I. For scheduling disciplines in the class we consider, the symmetry condition is necessary and sufficient if the service requirement distributions are different for different job classes or are non-exponential.*

(ii) *They have exponential service requirement distributions that do not depend on the job class. Here the scheduling discipline is arbitrary in the class of disciplines we consider. We call these stations type II.*

Furthermore, assume all blocking functions take the form:

$$b_{i\alpha}(\boldsymbol{k}_i) = h_{i\alpha}(k_{i\alpha})h_{i[\alpha]}(k_{i[\alpha]})h_i(k_i), \tag{22}$$

where $[\alpha]$ is the routing chain that contains job class α.

Then the equilibrium state probabilities have the product form

$$\pi(\boldsymbol{x}) = \frac{1}{G}\prod_i \Bigg[\prod_{1\leq l\leq k_i} \frac{h_i(l-1)}{f_i(l)} r_{i\kappa}(\sigma) \times \prod_{\Gamma} \prod_{1\leq l\leq k_{i\Gamma}} h_{i\Gamma}(l-1) \prod_{\alpha} \prod_{1\leq l\leq k_{i\alpha}} \frac{e_{i\alpha}h_{i\alpha}(l-1)}{\mu_{i\alpha}} \Bigg]. \tag{23}$$

Here i ranges over all stations, Γ ranges over all routing chains and α ranges over all job classes. Also, G is a normalization constant, selected such that the equilibrium state probabilities add up to one.

Proof. An elegant way of proving (23) is to guess the form of the reversed process and use this to verify the solution. Kelly [19] describes this method in detail. If we denote the quantities for the reversed process by primes, the method is based on the following relations [19, Theorem 1.13]:

$$\pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y}) = \pi(\boldsymbol{y})q'(\boldsymbol{y}, \boldsymbol{x}), \tag{24}$$

$$q'(\boldsymbol{x}) = q(\boldsymbol{x}). \tag{25}$$

The equilibrium distributions for both processes, the original and the reversed one, are the same.

In this case, the reversed process is almost the same network with rejection blocking. The only difference is that the scheduling disciplines are different:

$$\phi'_i(l, k_i) = \psi_i(l, k_i - 1), \tag{26}$$

$$\psi'_i(l, k_i) = \phi_i(l, k_i + 1). \tag{27}$$

Also, in the reversed network the state keeps track of the phases of service completed, not of the phases yet to be completed as in the original network. As a result, the expressions for the transition rates are somewhat more complex than (17) to (21), since we need to distinguish between the case in which the present phase is the last one, and the job leaves the station, or service has not yet finished, and the job advances to its next phase. Assuming that the job is in station i in class α and has reached phase s of its service, we need

$$\text{Pr(this is the last phase of service)} = \frac{g_{i\alpha;s}}{\sum_{t\geq s} g_{i\alpha;t}} = \frac{\nu_{i\alpha} g_{i\alpha;s}}{\mu_{i\alpha} r_{i\alpha}(s)}, \tag{28}$$

$$\text{Pr(this is not the last phase of service)} = \frac{\sum_{t\geq s+1} g_{i\alpha;t}}{\sum_{t\geq s} g_{i\alpha;t}} = \frac{r_{i\alpha}(s+1)}{r_{i\alpha}(s)}. \tag{29}$$

Using the equilibrium state distribution (23) and relation (24) we can compute the transition rates of the reversed process. With (28) and (29) we can interpret the result as the transition rates of another network.

In detail, we obtain for jobs arriving from outside in the original network to go to position m of station j, in class β with s phases of service left

$$q'(I_{jm;\beta s}(\boldsymbol{x}), \boldsymbol{x}) = p_{0,j\beta} \gamma b_{j\beta}(\boldsymbol{k}_j) \psi_j(m, k_j) g_{j\beta;s} \frac{\mu_{j\beta} f_j(k_j+1)}{e_{j\beta} b_{j\beta}(\boldsymbol{k}_j) r_{j\beta}(s)}$$

$$= p_{j\beta,0} \nu_{j\beta} f_j(k_j+1) \psi_j(m, k_j) \frac{\mu_{j\beta} g_{j\beta;s}}{\nu_{j\beta} r_{i\beta}(s)}. \tag{30}$$

In the reversed network, this corresponds to a job of class β that leaves the network from position m of station j after s phases of service.

For a job that leaves the original network from position l in station i we get

$$q'(D_{il}(\boldsymbol{x}), \boldsymbol{x}) = p_{i\kappa,0} \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i) \frac{e_{i\kappa} b_{i\kappa}(\boldsymbol{k}_i - \boldsymbol{u}_\kappa) r_{i\kappa}(1)}{\mu_{i\kappa} f(k_i)}$$

$$= p_{0,i\kappa} \gamma b_{i\kappa}(\boldsymbol{k}_i - \boldsymbol{u}_\kappa) \phi_i(l, k_i). \tag{31}$$

To derive (31) we used the identity

$$r_{i\alpha}(1) = \mu_{i\alpha} / \nu_{i\alpha}. \tag{32}$$

Equation (31) corresponds to a job that arrives from outside at station i, position l in class κ in the reversed network.

For a job in station i position l that advances to its next phase of service in the original network we have

$$q'(A_{il}(\boldsymbol{x}), \boldsymbol{x}) = \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i) \frac{r_{ik}(\sigma+1)}{r_{i\kappa}(\sigma)}. \tag{33}$$

This corresponds to the same thing in the reversed network, only that in the original network the job advances towards phase 1 while it advances to higher phases in the reversed network.

For a job that in the original network leaves position l of station i and goes to position m of station j in class β with s phases of service left we have

$$q'(T_{il,jm;\beta s}(\boldsymbol{x}), \boldsymbol{x}) = p_{i\kappa,j\beta}\,\nu_{i\kappa} f_i(k_i) b_{j\beta}(\boldsymbol{k}_j)\psi_j(m, k_j) g_{j\beta;s}$$

$$\times \frac{e_{i\kappa} b_{i\kappa}(\boldsymbol{k}_i - \boldsymbol{u}_\kappa) r_{i\kappa}(1)}{\mu_{i\kappa} f_i(k_i)} \frac{\mu_{j\beta} f_j(k_j)}{e_{j\beta} b_{j\beta}(\boldsymbol{k}_j) r_{j\beta}(s)},$$

$$= p_{j\beta,i\kappa}\,\nu_{j\beta} f_j(k_j) b_{i\kappa}(\boldsymbol{k}_i - \boldsymbol{u}_\kappa) \frac{\mu_{j\beta} g_{j\beta;s}}{\nu_{j\beta} r_{j\beta}(s)}. \tag{34}$$

In the reversed network this corresponds to a job of class β leaving position m of station j to go to position l of station i in class κ.

Finally, for a job that in the original network tried to go from position l of station i to station j in class β, but was rejected and returned to position n of station i we have

$$q'(T_{il,in;\kappa s}(\boldsymbol{x}), \boldsymbol{x}) = \nu_{i\kappa} f_i(k_i)\phi_i(l, k_i)\psi_i(n, k_i - 1) g_{i\kappa;s} \frac{r_{i\kappa}(1)}{r_{i\kappa}(s)}$$

$$\times \sum_{j\beta} p_{i\kappa,j\beta}(1 - b_{j\beta}(\boldsymbol{k}_j)),$$

$$= \nu_{i\kappa} f_i(k_i)\phi_i(l, k_i)\psi_i(n, k_i - 1) \frac{\mu_{i\kappa} g_{i\kappa;s}}{\nu_{i\kappa} r_{i\kappa}(s)}$$

$$\times \sum_{j\beta} p_{i\kappa,j\beta}(1 - b_{j\beta}(\boldsymbol{k}_j)). \tag{35}$$

In the reversed network, the job in position n of station i tried to leave station i to go to some other station j and class β but was rejected and returned to station i, position l.

The above transition rates for the reversed process were obtained from (17) through (21) by using the equilibrium state distribution (23) and relation (24). To complete the proof, we need to show that relation (25) also holds. We will consider a simpler (and more detailed) version of (25). It is clear that if we consider only changes at a particular position in a station, and for them an analogue of (25) holds, then (25) also holds by adding the result over all positions in all stations. This is the same idea that gives rise to the job local balance equations [8, 18, 19]. In the same way, by considering one job class and a single station one gets an analogue to the local balance equations [9].

In this case, the flow out of state $\boldsymbol{x}$ due to changes at position l of station i in the original network is given by

$$q_{il}(\boldsymbol{x}) = \nu_{i\kappa} f_i(k_i)\phi_i(l, k_i). \tag{36}$$

This follows since the later history of the job is irrelevant. The counterpart to (36) for the reversed process is obtained by adding up the transition rates (17) to (21).

The result is

$$q'_{il}(x) = \frac{f_i(k_i)}{r_{i\kappa}(\sigma)} [\mu_{i\kappa}\psi_i(l, k_i - 1) g_{i\kappa;\sigma} + \nu_{i\kappa}\phi_i(l, k_i) r_{i\kappa}(\sigma + 1)]. \tag{37}$$

Now we can specialize the results to stations of type I and II.

Station type I. It is seen that (25) is not automatically satisfied. Summing (37) over all l does not help, because of the $\kappa = \kappa_{il}$ and $\sigma = \sigma_{il}$ that appear in it. So one cannot get a sum of the ψ_i and ϕ_i alone as would be needed. On the other hand, if the scheduling discipline is symmetric we have, by definition (11),

$$\psi_i(l, k_i - 1) = f_i(l, k_i) \tag{38}$$

in which case (37) reduces to

$$q'_{il}(x) = \nu_{i\kappa} f_i(k_i) \phi_i(l, k_i). \tag{39}$$

We used the identity

$$\begin{aligned} \mu_{i\alpha} g_{i\alpha;s} + \nu_{i\alpha} r_{i\alpha}(s+1) &= \mu_{i\alpha}\left(g_{i\alpha;s} + \sum_{t \geqslant s+1} g_{i\alpha;t}\right) \\ &= \nu_{i\alpha} r_{i\alpha}(s). \end{aligned} \tag{40}$$

Comparing (39) with (36) shows that expressions (23) for $\pi(x)$ and expressions (30), (31), (33) and (34) do satisfy relations (24) and (25). So they are the equilibrium state distribution and the transition rates of the reversed process, respectively, as claimed. Moreover, if the service requirement distribution at station i is not exponential, or the service requirement distributions for different job classes are different, expression (23) is a solution iff the scheduling discipline is symmetric. This proves the claim for station type I.

Station type II. For stations of type II the above does not hold. But the analogue of local balance does. We have $\sigma = 1$ and

$$r_{i\kappa}(\sigma) = 1, \qquad r_{i\kappa}(\sigma + 1) = 0, \tag{41}$$

because there is only one phase of service.

Moreover, we have $g_{i\kappa;1} = 1$ and zero everywhere else. Also, we can write μ_i both for all $\mu_{i\alpha}$ and all $\nu_{i\alpha}$, there being only one phase and all service requirement distributions having the same mean.

Summing (36) over all positions in station i's queue we get

$$\begin{aligned} \sum_{1 \leqslant l \leqslant k_i} \mu_i f_i(k_i) \phi_i(l, k_i) &= \mu_i f_i(k_i) \sum_{1 \leqslant l \leqslant k_i} \xi_i(l, k_i) \\ &= \mu_i f_i(k_i). \end{aligned} \tag{42}$$

Note that we are able to simplify (42) only because all the $\nu_{i\alpha}$ are equal to μ_i.

The counterpart to (37) summed over all positions in station i is

$$f_i(k_i) \sum_{1 \le l \le k_i} \mu_i \psi_i(l, k_i - 1) = \mu_i f_i(k_i) \sum_{1 \le l \le k_i} \psi_i(l, k_i - 1)$$
$$= \mu_i f_i(k_i), \tag{43}$$

since the $g_{i\kappa;\sigma} = 1$ and the $r_{il}(\sigma) = 1$ disappear. The terms with ϕ_i appear multiplied by $r_{ik}(\sigma + 1) = 0$. So the expressions given satisfy both (24) and (25), as claimed. □

As pointed out when we described single stations, more general blocking functions can be considered. For definiteness, the proof is carried through for the case in which there is only one partition (except for single classes and all jobs) and we selected the routing chains as partition since we believe that this is the case of most practical interest. In general, however, one can consider several partitions of the job classes and the partitions may even be different for depending on the station.

Our equations (36) and (37) only consider one station and a routing chain at a time. We use reversible routing to derive (37) from (23). Under the hypothesis of Theorem 1, it is well known that there is a product form solution for classical networks without restrictions on the routing matrix [4, 18]. This can be proved in the same way as Theorem 1 [18, 19]. The restrictions are then needed only for routing chains in which there is blocking and only for flows to/from stations that block.

5. Distributions of occupancies and populations

As for classical networks, the form of the equilibrium state distribution (23) has interesting consequences.

For later convenience, we define the auxiliary functions

$$P_i(\boldsymbol{k}_i) = \prod_{1 \le l \le k_i} \frac{h_i(l-1)}{f_i(l)} \prod_{l'} \prod_{1 \le l \le k_{il'}} h_{il'}(l-1) \prod_{\alpha} \prod_{1 \le l \le k_{i\alpha}} \frac{e_{i\alpha} h_{i\alpha}(l-1)}{\mu_{i\alpha}}, \tag{44}$$

$$A_i(\boldsymbol{k}_i) = \binom{k_i}{k_{i1} k_{i2} \cdots k_{iC}} P_i(\boldsymbol{k}_i). \tag{45}$$

Corollary 2. *The equilibrium distribution of the occupancy depends only on the means of the service requirement distrbutions (insensitivity). The equilibrium occupancy distribution is given by*

$$\pi(\boldsymbol{n}) = \frac{1}{G} \prod_i P_i(\boldsymbol{k}_i) \tag{46}$$

and for populations the equilibrium distribution is given by

$$\pi(\boldsymbol{k}) = \frac{1}{G} \prod_i A_i(\boldsymbol{k}_i). \tag{47}$$

Here G is the same normalization constant as in the equilibrium state probabilities (23). *The functions P_i and A_i are defined by* (44) *and* (45) *respectively.*

Proof. To obtain the probability of the occupancy $\boldsymbol{n}$ all that has to be done is to sum the equilibrium state probabilities (23) over all possible number of phases of service left for each job. The form of the equilibrium state probabilities shows that we need sums of the form

$$\sum_s r_{i\alpha}(s) = \frac{\mu_{i\alpha}}{\nu_{i\alpha}} \sum_s \sum_{t \geq s} g_{i\alpha;t}$$
$$= \mu_{i\alpha} \sum_s \frac{s g_{i\alpha;s}}{\nu_{i\alpha}}$$
$$= 1. \tag{48}$$

Using (48) and the equilibrium state distribution (23) we get the equilibrium occupancy distribution (46). The service requirement distributions enter into the equilibrium occupancy distribution only by their means, as claimed.

The probability of an occupancy given by (46) does not depend on the order of the jobs in the stations. There are

$$\binom{k_i}{k_{i1} k_{i2} \cdots k_{iC}} \tag{49}$$

occupancies of station i that have the same population, thus proving the distribution for populations (47). □

Open networks have particularly simple equilibrium distributions as stated in the following corollary.

Corollary 3. *In an open network, the states of the stations are independent. The same holds for the occupancies and the populations*

Proof. This is an immediate consequence of the product form of the equilibrium state, occupancy and population distributions for open networks, equations (23), (46) and (47), respectively. □

Remarks

Theorem 1 was originally conjectured based on the results of Cohen [11], and Van Dijk and Tijms [13], who proved product form for cyclic networks with two stations of types I and II. Class changes are not allowed in these models. The routing is reversible in cyclic networks with two stations, and Hordijk and Van Dijk [15] proved that in closed networks with a single job class, exponential service requirement distributions and reversible routing, the equilibrium state probabilities have product form. Hordijk and van Dijk [24, 26] studied similar models like in our case.

However, they do not allow job class changes in the network. The distributions are assumed to be exponential. The blocking protocol is different. Moreover, they assume that the product form solution exists. The proof is then carried out based on that assumed solution. Von Brand [5] proved Theorem 1 by a different method. He also investigated several other related models with rejection blocking and derived exact algorithms to compute performance measures. Pittel [29] proved a similar result for multiple job classes when the scheduling discipline is processor sharing. Another hint was the result of Melamed [21], who proved that *classical* (non-blocking) networks with reversible routing are reversible [19] if the service requirement distributions are exponential and the scheduling disciplines are symmetric. From Melamed's [21] result and the truncation theorem for reversible Markov processes [19, Lemma 1.9] it would follow by a messy induction argument that networks similar to these but with rejection blocking also have product form solutions, and that the solution is precisely of the form (23). Class changes were also included, since in classical networks class changes can be allowed [18]. As can be seen, there are very strong similarities between classical networks and networks with rejection blocking and reversible routing.

6. The departure processes

Using the proof of Theorem 1 we can deduce some further properties of the network in a simple way. The fact that the reversed process is very similar to the original processs is of great help in this.

Corollary 4. *The streams of jobs of each class that leave the network (either after traversing it or after being rejected when trying to enter) are independent Poisson streams.*

Proof. In the original description of the network we assume a single Poisson arrival stream that is split by the $p_{0,j\beta}$. So the arrivals of each class form independent Poisson streams. In the proof of Theorem 1 we found that the reversed process is of the same type, i.e. a network with independent Poisson arrivals for each job class. Now each arrival in the reversed process corresponds to a departure or a rejection in the original process, and the result follows. □

Corollary 5. *The distribution of states at the instants at which jobs of any particular class arrive at the network is the equilibrium state distribution. The same holds for the distribution of the states at instants at which jobs depart from the network, either after traversing the network or after being rejected on arrival.*

Proof. The arrival process for jobs of any particular class is Poisson. To check the state of the network at arrival instants is then the same as checking it at random, and the first claim follows. Departures correspond to arrivals in the reversed process. So the second assertion follows by the same argument and the fact that the equilibrium state distributions of the original process and its reverse are the same. □

For the following corollary we need the definition of quasi-reversible process [19]. A process in equilibrium is called *quasi-reversible* if

(i) the state of the network at time t is independent of arrivals after t;

(ii) the state of the network at time t is independent from departures prior to t.

Corollary 6. *The open and mixed networks described in Section* 3 *are quasi-reversible.*

Proof. We need to check the two conditions of the above definition. But (i) is obvious from the definition of the network, and similarly (ii) is clear by considering the reversed process. □

7. Conclusions

We show that a class of queueing networks with rejection blocking has a product form equilibrium state distribution, and that the distribution of the population is insensitive. The results are strikingly similar to the corresponding results for classical (non-blocking) networks. This poses the question of how far the similarities go. For example, one might expect that there is a simple relation between the state of the network at equilibrium and the state of the network at the instants at which jobs arrive at a station. For classical networks, this is called the Arrival Instant Distribution Theorem [20, 30]. In the case of classical networks, all jobs that arrive at a station are accepted. In the models considered here this is not necessarily so, and one could also consider an analogous Acceptance Instant Distribution Theorem for rejection blocking networks.

References

[1] I.F. Akyildiz and H. von Brand, Duality in open and closed Markovian queueing networks with rejection blocking, Computer Science Technical Report TR 87-011, Louisiana State University, Baton Rouge, LA, April 1987.

[2] S. Balsamo and G. Iazeolla, Some equivalence properties for queueing networks with and without blocking, in: A.K. Agrawal and S. Tripathi, eds., *Performance '83* (North-Holland, New York, 1983) 351-360.

[3] A.D. Barbour, Networks of queues and the method of stages, *Adv. Appl. Probability* **8**(3) (1976) 584-591.

[4] F. Baskett, K.M. Chandy, R.R. Muntz and F.G. Palacios, Open, closed and mixed networks of queues with different classes of customers, *J. ACM* **22**(2) (1975) 249-260.

[5] H. von Brand, Queueing networks with blocking, Ph.D. Thesis, Louisiana State University, Baton Rouge, LA, July 1987.

[6] J.P. Buzen, Computational algorithms for closed queueing networks with exponential servers, *Com. ACM* **16**(9) (1973) 335-359.

[7] P.J. Caseau and G. Pujolle, Throughput capacity of a sequence of queues with blocking due to finite waiting room, *IEEE Trans. Software Engineering* **SE-5**(6) (1979) 631-642.

[8] K.M. Chandy and A.J. Martin, A characterization of product-form queueing networks, *J. ACM* **30**(2) (1983) 286-299.

[9] K.M. Chandy, J.H. Howard, Jr and D.F. Towsley, Product form and local balance in queueing networks, *J. ACM* **24**(2) (1977) 250-263.

[10] K.M. Chandy and C.H. Sauer, Computational algorithms for product form queueing networks, *Com. ACM* **23**(10) (1980) 573-583.

[11] J.W. Cohen, The multiple phase service network with generalized processor sharing, *Acta Informatica* **12** (1979) 245-284.

[12] D.R. Cox, A use of complex probabilities in the theory of stochastic processes, *Proc. Camb. Phil. Soc.* **51** (1955) 313-319.

[13] N.M. van Dijk and H.C. Tijms, Insensitivity in two-node blocking models with applications, in: O.J. Boxma, J.W. Cohen and H.C. Tijms, eds., *Teletraffic Analysis and Computer Performance Evaluation* (Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986) 329-340.

[14] W.J. Gordon and G.F. Newell, Closed queueing systems with exponential servers, *Oper. Res.* **15**(2) (1967) 254-265.

[15] A. Hordijk and N. van Dijk, Networks of queues with blocking, in: F.J. Klystra, ed., *Performance '81* (North-Holland, New York, 1981) 51-65.

[16] J.R. Jackson, Jobshop-like queueing systems, *Management Sci.* **10**(1) (1963) 131-142.

[17] F.P. Kelly, Networks of queues with customers of different types, *J. Appl. Probability* **12** (1975) 542-554.

[18] F.P. Kelly, Networks of queues, *Adv. Appl. Probability* **8**(2) (1976) 416-432.

[19] F.P. Kelly, *Reversibility and Stochastic Networks* (Wiley, Chichester, England, 1979).

[20] S.S. Lavenberg and M. Reiser, Stationary state probabilities at arrival instants for closed queueing networks with multiple types of customers, *J. Appl. Probability* **17**(4) (1980) 1048-1061.

[21] B. Melamed, On the reversibility of queueing systems, *Stochastic Processes and their Applications* **13** (1982) 227-234.

[22] R.R. Muntz and J.W.-N. Wong, Efficient computational procedures for closed queueing network models, in: *Proc. 7th Hawaii Int. Conf. on System Sciences* (1974) 33-36.

[23] A.S. Noetzel, A generalized queueing discipline for product form network solutions, *J. ACM* **26**(4) (1979) 779-793.

[24] A. Hordijk and N.M. van Dijk, Adjoint processes, job-local-balance and insensitivity of stochastic networks, *Bull. 44th Session Internat. Stat. Inst* **50** (1982) 776-788.

[25] R.O. Onvural and H.G. Perros, On equivalences of blocking mechanisms in queueing networks with blocking, *Oper. Res. Lett.* **5**(5) (1986).

[26] A. Hordijk and N.M. van Dijk, Networks of queues, Part I: Job-local-balance and the adjoint process; Part II: General routing and service characteristics, in: *Proc. Internat. Conf. on Modeling for Computer Systems* (Springer, Berlin, 1983) 158-205.

[27] M. Reiser and H. Kobayashi, Queueing networks with multiple closed chains: Theory and computational algorithms, *IBM J. Res. Dev.* **19** (1975) 283-294.

[28] B. Pittel, Closed exponential networks with queues with saturation: The Jackson type stationary distribution and its asymptotic analysis, *Math. Oper. Res.* **4**(4) (1979) 367-378.

[29] M. Reiser and S.S. Lavenberg, Mean value analysis of closed multichain queueing networks, *J. ACM* **27**(2), (1980) 313-322; Corrigendum **28**(3), (1981) 629.

[30] K.C. Sevcik and I. Mitrani, The distribution of queueing network states at input and output instants, *J. ACM* **28**(2), (1981) 358-371.

[31] J.R. Spirn, Queueing networks with random selection for service, *IEEE Trans. Software Engineering* **SE-5**(3) (1979) 287-289.

[32] H.C. Tijms, *Stochastic Modeling and Analysis: A Computational Approach* (Wiley, Chichester, 1986).